Algorithms based on sparsity hypotheses for robust estimation of the noise standard deviation in presence of signals with unknown distributions and concurrences

Dominique Pastor, François-Xavier Socheleau

Collection des rapports de recherche
de Télécom Bretagne

RR-2010001-SC

Algorithms based on sparsity hypotheses for robust estimation of the noise standard deviation in presence of signals with unknown distributions and occurrences

Dominique Pastor (Télécom Bretagne, Lab-STICC)
François-Xavier Socheleau (Télécom Bretagne, Lab-STICC)

Abstract

In many applications, d-dimensional observations result from the random presence or absence of random signals in independent and additive white Gaussian noise. An estimate of the noise standard deviation can then be very useful to detect or to estimate these signals, especially when standard likelihood theory cannot apply because of too little prior knowledge about the signal probability distributions. Recent results and algorithms have then emphasized the interest of sparsity hypotheses to design robust estimators of the noise standard deviation when signals have unknown distributions. As a continuation, the present paper introduces a new robust estimator for signals with probabilities of presence less than or equal to one half. In contrast to the standard MAD estimator, it applies whatever the value of d. This algorithm is applied to image denoising by wavelet shrinkage as well as to non-cooperative detection of radiocommunications. In both cases, the estimator proposed in the present paper outperforms the standard solutions used in such applications and based on the MAD estimator. The Matlab code corresponding to the proposed estimator is available at http://perso.telecom-bretagne.eu/pastor.

Keywords

Sparsity, robust statistics, non parametric statistics, MAD estimation, wavelet shrinkage, soft thresholding, image denoising, Communication Electronic Support, cognitive radio.

Contents

1 Introduction

In many applications, observations are d-dimensional random vectors that result from the random presence of signals in independent and additive white Gaussian noise (AWGN). In most cases, two difficulties are met simultaneously to detect or to estimate the noisy signals. On the one hand, very little is generally known about the signals or about most of their describing parameters so that the probability distributions of the signals are partially or definitely unknown. On the other hand, the noise standard deviation is often unknown and must be estimated so as to process the observations.

Such situations where the distributions of the signals to detect are not fully known and the noise standard deviation must be estimated are frequently encountered in sonar and radar processing because the echoes, received by a sonar or a radar system, result from a convolution between a possibly known transmitted pulse and an often unknown environment. A second typical example, discussed in the experimental part of this paper, is that of spectrum sensing performed in a context of communication electronic support (CES) [1] or cognitive radio [2]. In such applications, it is desirable to perform the detection of some signals of interest based on an observation that is usually a noisy mixture of signals with unknown distributions and occurrences. Since very little is generally known about the transmitted signals, it is often relevant to estimate the noise standard deviation via robust scale estimators. Such estimators are said to be robust in the sense that they are not excessively affected by small departures from the assumed model. In our case, the model is the Gaussian distribution of the noise and the signals are considered as outliers with respect to this model. A widely used robust scale estimator is the median absolute deviation about the median (MAD) defined by [3, 4]

$$\mathrm{MAD} = b \times \mathrm{med}_i |Y_i - \mathrm{med}_j Y_j|, \tag{1}$$

where $Y = Y_1, \cdots, Y_m$ is the observation, b is a constant needed to make the estimator Fisher-consistent at the model distribution and $\mathrm{med}_i Y_i$ stands for the median value of the observation. For instance, $b \approx 1.4826$ to make the MAD consistent at the normal distribution. The MAD popularity comes from its high asymptotic breakdown point (50%) and its influence function bounded by the sharpest possible bound among all scale estimators [4]. It was also defined in [5] as "the single most useful ancillary estimate of scale" so that it is often used as an initial value for more efficient robust estimators such as M-estimators. Despite its high breakdown point, the performance of the MAD degrades significantly when the proportion of outliers increases. As an alternative to the MAD, we here address the problem of estimating the noise standard deviation in applications where the number or the amplitudes of the outliers are too large for the MAD estimator to perform well.

In a recent work [6, 7], as a continuation of [8] and [9], we presented an estimate, namely the MC-ESE (Modified Complex Essential Supremum Estimate), of the noise standard deviation with application to CES. The MC-ESE applies to observations where the signals have unknown probability distributions and unknown probabilities of presence less than or equal to one half in presence of independent AWGN. No prior knowledge on the signals is required. In fact, the MC-ESE is an heuristic algorithm whose theoretical background is [8, Theorem 1]. This theoretical result involves hypotheses that bound our lack of prior knowledge about the amplitudes and occurrences of the signals. These hypotheses are particularly suitable in case of observations returned by a sparse transform, that is, a transform representing a signal by coefficients that are mostly small except a few ones that have large amplitudes. In this sense, the short term Fourier transform (STFT) of a wideband signal is sparse in CES applications. In [6], we then applied the MC-ESE to the detection of non-cooperative communications in CES. It followed from the experimental results given in [6] that the MC-ESE outperforms the standard MAD estimator when the number of communications to detect and, thus, the number of outliers increases so as to approach 50 percent of the sample size. However, the MC-ESE raises one open question and has one drawback. The open question is that the performance of the MC-ESE is not predicted by [8, Theorem 1]. In fact, [8, Theorem 1] is a theoretical result concerning asymptotic situations where the number of observations is large and the amplitudes of the signals corrupted by independent AWGN are either big or small, whereas, surprisingly enough, the MC-ESE is capable of achieving a good estimate of the noise standard deviation whatever the amplitudes of the signals. The drawback is that the MC-ESE is computationally expensive because of the minimization routine it involves.

In this paper, we present a new type of noise standard deviation estimate, called fast-ESE for fast essential supremum estimate (ESE). The term ESE is used after the terminology of [8], which is the theoretical baseline of this work. First, the fast-ESE derives from a corollary of [8, Theorem 1] and is far more theoretically justified than the MC-ESE. Second, the fast-ESE computational cost is far lesser than that of the MC-ESE. Thence, its name. The fast-ESE belongs to the family of L-estimators that are by definition based on a linear combination of order statistics. In its general form, the fast-ESE requires prior knowledge of what we define as the minimum signal to noise ratio (SNR). According to some simulations on synthetic signals suitably chosen, it turns out that this parameter can be fixed, which leads to the universal fast-ESE. The universal fast-ESE can be regarded as an alternative to the MAD estimator since both need no prior knowledge about the signals and their distributions. This is the reason why we test the universal fast-ESE in two rather natural applications of such an estimator. First, in image denoising by wavelet shrinkage and as a continuation of [10, 11], the experimental results presented below show that the universal fast-ESE outperforms the MAD. In this application, a comparison between the universal fast-ESE and the MC-ESE is meaningless since the latter is designed for two-dimensional or complex observations whereas the application concerns wavelet coefficients, which are real values. The second application makes it possible to compare the universal fast-ESE, the MC-ESE and the MAD, since it concerns the detection of non-cooperative communications.

In the next section, we summarize and discuss our previous results on robust estimators, that is, [8, Theorem 1] and the MC-ESE. The fast-ESE is then introduced in section 3 after establishing the corollary of [8, Theorem 1] from which it derives. According to some simulations, we can fix the parameter on which the fast-ESE depends and derive the universal fast-ESE. The application of the universal fast-ESE to image denoising by wavelet shrinkage is treated in section 4.1, whereas its application to cognitive radio and CES is addressed in section 4.2. Despite the very good results achieved by the universal fast-ESE, some questions remain open about the fast-ESE and the universal fast-ESE. These questions concern the theoretical performance of such an estimator, the possibility to extend the concepts underlying the fast-ESE and the universal fast-ESE so as to propose a whole class of robust estimators, amongst which the MAD, since the fast-ESE and the universal fast-ESE are also based on order statistics. These remarks and the prospects that the fast-ESE and the universal fast-ESE open are discussed in the concluding part of the paper, namely, section 5.

2 Theoretical background

Signal processing applications often involve sequences of d-dimensional real random observations such that each observation is either the sum of some random signal and independent noise or noise alone. In most cases, noise is reasonably assumed to be “white and Gaussian” in that it is Gaussian distributed with null mean and covariance matrix proportional to the $d \times d$ identity matrix $\mathbf{I}_d$. Summarizing, we say that each observation results from some signal randomly present or absent in independent AWGN. For reasons described in the introduction, the problem is the estimation of the noise standard deviation in the general case where the probability distributions of the signals are unknown. In this respect, it is proved in [8] that, when the observations are independent and the probabilities of presence of the signals are upper-bounded by some value in [0,1), the noise standard deviation is the only positive real number satisfying a specific convergence criterion when the number of observations and the minimum amplitude of the signals tend to infinity. This convergence involves neither the probability distributions nor the probabilities of presence of the signals. Our new estimator presented in section 3, as well as the MC-ESE of [6] and [9], basically derive from this theoretical result that we must, therefore, recall properly. To this end, we need some material and hypotheses, used throughout. This material is presented in the next section. The theoretical result of [8] and on which the MC-ESE and fast-ESE are based is stated and commented in section 2.2. The MC-ESE is then briefly described in section 2.4 since the fast-ESE will be compared to it.

2.1 Preliminary material

To begin with, every random vector and every random variable encountered hereafter is assumed to be real-valued and defined on the same probability space $(\Omega, \mathscr{B}, \mathrm{P})$ for every $\omega \in \Omega$. As usual, if a property $\mathscr{P}$ holds true almost surely, we write $\mathscr{P}$ (a-s). Every random vector considered below is d-dimensional. The set of all d-dimensional real random vectors, that is, the set of all measurable maps of Ω into $\mathbb{R}^d$, is denoted by $\mathscr{M}(\Omega, \mathbb{R}^d)$ and the sequences of d-dimensional random vectors defined on Ω is denoted by $\mathscr{M}(\Omega, \mathbb{R}^d)^{\mathbb{N}}$. In what follows, $\|\cdot\|$ is the standard Euclidean norm in $\mathbb{R}^d$. For every given random vector $Y : \Omega \to \mathbb{R}^d$ and any $\tau \in \mathbb{R}$, the notation $\mathrm{I}(\|Y\| \leqslant \tau)$ stands for the indicator function of the event $\left[\|Y\| \leqslant \tau\right]$. If Y is any d-dimensional random vector (resp. any random variable), the probability that Y belongs to some Borel set A of $\mathbb{R}^d$ (resp. $\mathbb{R}$) is denoted by $\mathrm{P}[Y \in A]$.

Given some positive real number σ_0, we say that a sequence $X = (X_k)_{k \in \mathbb{N}}$ of d-dimensional real random vectors is a ***d-dimensional white Gaussian noise*** (WGN) with standard deviation σ_0 if the random vectors X_k, $k \in \mathbb{N}$, are independent and identically Gaussian distributed with null mean vector and covariance matrix $\sigma_0^2 \mathbf{I}_d$. The ***minimum amplitude*** of an element $\Lambda = (\Lambda_k)_{k \in \mathbb{N}}$ of $\mathscr{M}(\Omega, \mathbb{R}^d)^{\mathbb{N}}$ is defined as the supremum $\mathsf{a}(\Lambda)$ of the set of those $\alpha \in [0, \infty]$ such that, for every natural number k, $\|\Lambda_k\|$ is larger than or equal to α (a-s):

$$\mathsf{a}(\Lambda) = \sup\{\alpha \in [0, \infty] : \forall k \in \mathbb{N}, \|\Lambda_k\| \geq \alpha \text{ (a-s)}\}. \tag{2}$$

The minimum amplitude has some properties easy to verify. First, for every given $\Lambda = (\Lambda_k)_{k \in \mathbb{N}}$, $\mathsf{a}(\Lambda)$, $\|\Lambda_k\| \geq \mathsf{a}(\Lambda)$ for every $k \in \mathbb{N}$; second, given $\alpha \in [0, \infty]$, $\mathsf{a}(\Lambda) \geq \alpha$ if and only if, for every $k \in \mathbb{N}$, $\|\Lambda_k\| \geq \alpha$ (a-s). If f is some map of $\mathscr{M}(\Omega, \mathbb{R}^d)^{\mathbb{N}}$ into $\mathbb{R}$, we will then say that the limit of f is $\ell \in \mathbb{R}$ when $\mathsf{a}(\Lambda)$ tends to ∞ and write that $\lim_{\mathsf{a}(\Lambda) \to \infty} f(\Lambda) = \ell$ if

$$\lim_{\alpha \to \infty} \sup\left\{|f(\Lambda) - \ell| : \Lambda \in \mathscr{M}(\Omega, \mathbb{R}^d)^{\mathbb{N}}, \mathsf{a}(\Lambda) \geqslant \alpha\right\} = 0, \tag{3}$$

that is, if, for any positive real value η, there exists some $\alpha_0 \in (0,\infty)$ such that, for every $\alpha \geq \alpha_0$ and every $\Lambda \in \mathscr{M}(\Omega,\mathbb{R}^d)^{\mathbb{N}}$ such that $\mathsf{a}(\Lambda) \geq \alpha$, we have $|f(\Lambda) - \ell| \leqslant \eta$.

Given some non-negative real number a, $\mathrm{L}^a(\Omega,\mathbb{R}^d)$ stands for the set of those d-dimensional real random vectors $Y : \Omega \to \mathbb{R}^d$ for which $\mathbb{E}[\|Y\|^a] < \infty$. We hereafter deal with the set of those elements $\Lambda = (\Lambda_k)_{k\in\mathbb{N}}$ of $\mathscr{M}(\Omega,\mathbb{R}^d)^{\mathbb{N}}$ such that $\Lambda_k \in \mathrm{L}^a(\Omega,\mathbb{R}^d)$ for every $k \in \mathbb{N}$ and $\sup_{k\in\mathbb{N}} \mathbb{E}[\|\Lambda_k\|^a]$ is finite. This set is hereafter denoted $\ell^\infty(\mathbb{N},\mathrm{L}^a(\Omega,\mathbb{R}^d))$.

A ***thresholding function*** is any non-decreasing continuous and positive real function $\theta : [0,\infty) \to (0,\infty)$ such that

$$\theta(\rho) = C\rho + \gamma(\rho) \tag{4}$$

where $0 < C < 1$, $\gamma(\rho)$ is positive for sufficiently large values of ρ and $\lim_{\rho\to\infty} \gamma(\rho) = 0$. Given any $q \in [0,\infty)$, Υ_q will stand for the map of $[0,\infty)$ into $[0,\infty)$ defined for every $x \in [0,\infty)$ by

$$\Upsilon_q(x) = \int_0^x t^{q+d-1} e^{-t^2/2} \mathrm{d}t. \tag{5}$$

2.2 A limit theorem

On the basis of the material proposed above, it is easy to model a sequence of observations where random signals are either present or absent in independent AWGN with standard deviation σ_0. It suffices to consider a sequence $Y = (Y_k)_{k\in\mathbb{N}}$ such that, for every $k \in \mathbb{N}$, $Y_k = \varepsilon_k \Lambda_k + X_k$ where $\varepsilon = (\varepsilon_k)_{k\in\mathbb{N}}$ is a sequence of random variables valued in $\{0,1\}$, $\Lambda = (\Lambda_k)_{k\in\mathbb{N}}$ stands for some sequence of d-dimensional real random vectors and $X = (X_k)_{k\in\mathbb{N}}$ is some d-dimensional WGN with standard deviation σ_0. We write $Y = \varepsilon\Lambda + X$ and, in this model, Y is the sequence of observations, Λ, the sequence of signals that can be observed in independent AWGN represented by X and ε_k models the possible occurrence of Λ_k so that each Y_k obeys a binary hypothesis testing model where the null hypothesis is that $\varepsilon_k = 0$ and the alternative one is that $\varepsilon_k = 1$. We can now state the following result, established in [8].

Theorem 1 *Let $Y = (Y_k)_{k\in\mathbb{N}}$ be some element of $\mathscr{M}(\Omega,\mathbb{R}^d)^{\mathbb{N}}$ such that $Y = \varepsilon\Lambda + X$ where $\Lambda = (\Lambda_k)_{k\in\mathbb{N}}$, $X = (X_k)_{k\in\mathbb{N}}$ and $\varepsilon = (\varepsilon_k)_{k\in\mathbb{N}}$ are an element of $\mathscr{M}(\Omega,\mathbb{R}^d)^{\mathbb{N}}$, some d-dimensional WGN with standard deviation σ_0 and a sequence of random variables valued in* $\{0,1\}$ *respectively.*

Assume that

(A1) *for every $k \in \mathbb{N}$, Λ_k, X_k and ε_k are independent;*

(A2) *the random vectors Y_k, $k \in \mathbb{N}$, are independent;*

(A3) *the set of priors $\{\mathrm{P}[\varepsilon_k = 1] : k \in \mathbb{N}\}$ has an an upper bound $\mathrm{p} \in [0,1)$ and the random variables ε_k, $k \in \mathbb{N}$, are independent;*

(A4) *there exists some $\nu \in (0,\infty]$ such that $\Lambda \in \ell^\infty(\mathbb{N}, \mathrm{L}^\nu(\Omega,\mathbb{R}^d))$.*

Let r and s be any two non-negative real numbers such that $0 \leqslant s < r \leqslant \nu/2$. Given some natural number m and any pair (σ, T) of positive real numbers, define the random variable $\Delta_m(\sigma,T)$ by

$$\Delta_m(\sigma,T) = \left| \frac{\sum_{k=1}^{m} \|Y_k\|^r \mathrm{I}(\|Y_k\| \leqslant \sigma T)}{\sum_{k=1}^{m} \|Y_k\|^s \mathrm{I}(\|Y_k\| \leqslant \sigma T)} - \sigma^{r-s} \frac{\Upsilon_r(T)}{\Upsilon_s(T)} \right|. \tag{6}$$

Then, given any thresholding function θ, σ_0 is the unique positive real number σ such that, for every $\beta_0 \in (0,1]$,

$$\lim_{\mathrm{a}(\Lambda)\to\infty} \left\| \limsup_{m\to\infty} \Delta_m(\sigma, \beta\theta(\mathrm{a}(\Lambda)/\sigma)) \right\|_\infty = 0 \tag{7}$$

uniformly in $\beta \in [\beta_0, 1]$.

PROOF: see [8, Appendix A] ■

We recall that, given a sequence $(u_q)_{q\in\mathbb{N}}$ of real numbers, the upper limit $\limsup_q u_q$ of this sequence is the limit, when q tends to ∞ of the non-increasing sequence $\sup\{u_k : k \geqslant q\}$. We also remind the reader that, given a complex random variable Z, $\|Z\|_\infty$ is called the essential supremum of Z and is defined by $\|Z\|_\infty = \inf\{\rho \in [0,\infty) : |Z| \leqslant \rho \text{ (a-s)}\}$ so that $|Z| \leqslant \rho$ (a-s) if and only if $\rho \geqslant \|Z\|_\infty$.

At this stage, some comments are required. To begin with, note that assumption **(A3)** can be regarded as a weak assumption of sparsity because it imposes that the signals are not always present without imposing small probabilities of occurrence for the signals. Our second remark is that the result of theorem 1 can be seen as a method of moments, as illustrated by the heuristic approach given in [7] and [6]. In short, this heuristic approach considers the sample moments

$$\mu_\gamma(q,\tau) = \frac{1}{q} \sum_{k=1}^{q} \|Y_k\|^\gamma \mathrm{I}(\|Y_k\| \leqslant \sigma_0 \tau), \gamma \in [0,\infty) \tag{8}$$

for any $\tau \in [0,\infty)$ and any $q \in \mathbb{N}$. If the random vectors Λ_k are independent and identically distributed (iid), the strong law of large numbers makes it possible to justify the approximation

$$\mu_r(q,\tau)/\mu_s(q,\tau) \approx \mathbb{E}\left[\|Y_k\|^r \mathrm{I}(\|Y_k\| \leqslant \sigma_0\tau)\right]/\mathbb{E}\left[\|Y_k\|^s \mathrm{I}(\|Y_k\| \leqslant \sigma_0\tau)\right]. \tag{9}$$

On the second hand, when each $\|\Lambda_k\|$ is large enough with respect to σ_0, we can expect the existence of a suitable threshold height τ capable of distinguishing observations where the signals are present from observations where noise only is present, so that we can approximatively write, thanks to this threshold, that

$$\mathbb{E}\left[\|Y_k\|^\gamma \mathrm{I}(\|Y_k\| \leqslant \sigma_0\tau)\right] \approx \mathbb{E}\left[\|X_k\|^\gamma \mathrm{I}(\|X_k\| \leqslant \sigma_0\tau)\right] \mathrm{P}[\varepsilon_k = 0] \tag{10}$$

for any $\gamma \in [0,\infty)$. By combining this approximation with Eq. (9), we conclude that, in a certain sense to specify, $\mu_r(q,\tau)/\mu_s(q,\tau)$ tends to

$$\mathbb{E}\left[\|X_k\|^r \mathrm{I}(\|X_k\| \leqslant \sigma_0\tau)\right]/\mathbb{E}\left[\|X_k\|^s \mathrm{I}(\|X_k\| \leqslant \sigma_0\tau)\right] = \sigma_0^{r-s}\Upsilon_r(\tau)/\Upsilon_s(\tau), \tag{11}$$

when q and the norms of the signals are large enough and where the right-hand-side of Eq. (11) derives from the fact that the $\|X_k\|^2$ has the centered chi-2 distribution with d degrees of freedom. The difficulty in combining Eqs. (9) and (11) relies on the fact that the former involves an almost sure convergence when q tends to infinity whereas the latter involves a convergence when the norms of the signals are large enough. Therefore, the resulting convergence criterion, according to which $\mu_r(q,\tau)/\mu_s(q,\tau)$ tends to $\sigma_0^{r-s}\Upsilon_r(\tau)/\Upsilon_s(\tau)$ when q and the norms of the signals are large enough, cannot be trivial. It is a surprising and unforeseen fact that this convergence is satisfied even when the signals are not iid, whatever their probability distributions, if we assume **(A3)** and **(A4)** and thanks to the notions of minimum amplitude and thresholding function. Even though the existence of some convergence can somewhat be guessed, the specific form of the convergence criterion of Eq. (7) is not intuitive.

2.3 The Essential Supremum Estimate (ESE)

With the same notations and assumptions of theorem 1, let $Y_1,\ldots,Y_m$ be m observations and assume that the minimum amplitude of the signals Λ_k for $k = 1,\ldots,m$ is a. The discussion in [8] about Eq. (7) suggests estimating σ_0 by

$$\breve{\sigma}_0 = \operatorname*{argmin}_{\sigma} \sup_{\ell \in \{1,\ldots,L\}} \Delta_m(\sigma, \beta_\ell \theta(\mathsf{a}/\sigma)) \tag{12}$$

where $L \in \mathbb{N}$ and $\beta_\ell = \ell/L$ for every $\ell \in \{1,\ldots,L\}$. Because of the role played by the essential supremum in Eq. (7), the estimate $\breve{\sigma}_0$ is called the Essential Supremum Estimate (ESE) of the noise standard deviation.

For practical usage, we must choose values for r and s, an appropriate threshold function θ and a search interval. To this end, we restrict our attention to particular cases of immediate and practical interest. These cases concern assumptions **(A3)** and **(A4)**. First, in practice, signals have finite energy, which amounts to assuming **(A4)** with $\nu = 2$. Moreover, in many signal processing applications, signals of interest are less present than absent, so that we can reasonably restrict our attention to the case where **(A3)** is satisfied with $\mathrm{p} = 1/2$. It is worth noticing that, when the amplitudes of the signals are large enough, we can say that the sequence of observations $(Y_k)_{k\in\mathbb{N}}$ is sparse in the sense that most of the random vectors Y_k are due to noise alone and at most half of them contain signals.

Since $\nu = 2$, we choose $s = 0$ and $r = 1$ in Eq. (12). Other choices could be made but have not been studied, yet. The choice of the thresholding function relates now to the assumption $\mathrm{p} = 1/2$. Specifically, for every given $\rho \in [0,\infty)$, let $\xi(\rho)$ be the unique positive solution for x in the equation

$$
{}_0\mathrm{F}_1(\mathrm{d}/2; \rho^2 x^2/4) = e^{\rho^2/2}, \tag{13}
$$

where ${}_0\mathrm{F}_1$ is the generalized hypergeometric function [12, p. 275]. This map ξ is a thresholding function with $C = 1/2$ [8]. Under assumption **(A3)** with $\mathrm{p} = 1/2$, that is, when signals are less present than absent, ξ is particularly relevant because it follows from [13, Theorem VII-1] that the thresholding test with threshold height $\sigma_0\xi(\mathsf{a}/\sigma_0)$ is capable of making a decision on the value of any ε_k with a probability of error that decreases rapidly with a/σ_0. By thresholding test with threshold height $\sigma_0\xi(\mathsf{a}/\sigma_0)$, we mean the test that returns 1 when the norm of any given observation exceeds $\sigma_0\xi(\mathsf{a}/\sigma_0)$ and 0 otherwise. The use of the thresholding function ξ thus optimizes the approximation of Eq. (10) and, therefore, favors the convergence in Eq. (7). With this choice for the thresholding function, an appropriate search interval $[\sigma_{\min}, \sigma_{\max}]$ for σ in Eq. (12) can be calculated. Denoting by $Y_{[k]}$, $k = 1, 2, \ldots, m$, the sequence of observations $Y_1, Y_2, \ldots, Y_m$, sorted by increasing norm, the lower bound of the search interval is $\sigma_{\min} = \| Y_{[k_{\min}]} \| / \sqrt{d}$ where

$$
k_{\min} = m/2 - hm, \tag{14}
$$

$h = 1/\sqrt{4m(1-Q)}$ and Q is a positive real number less than or equal to $1 - \frac{m}{4(m/2-1)^2}$. The upper bound of the search interval is fixed to $\sigma_{\max} = \| Y_{[m]} \| / \sqrt{d}$. The details of the construction are given in [8, Section 3.2].

With the parameters set as above, the ESE associated with the thresholding function ξ can be computed as:

$$\breve{\sigma}_0 = \underset{\sigma \in [\sigma_{\min}, \sigma_{\max}]}{\operatorname{argmin}} \sup_{\ell \in \{1,\ldots,L\}} \left\{ \left| \frac{\sum_{k=1}^{m} \| Y_k \| \mathrm{I}(\| Y_k \| \leqslant \beta_\ell \sigma \xi(\mathsf{a}/\sigma))}{\sum_{k=1}^{m} \mathrm{I}(\| Y_k \| \leqslant \beta_\ell \sigma \xi(\mathsf{a}/\sigma))} - \sigma \frac{\Upsilon_1(\beta_\ell \xi(\mathsf{a}/\sigma))}{\Upsilon_0(\beta_\ell \xi(\mathsf{a}/\sigma))} \right| \right\}. \tag{15}$$

2.4 The Modified Complex Essential Supremum Estimate

In general, the minimum amplitude of the signals is unknown in practice. This is why we introduced the MC-ESE in [9] and [6]. This estimator requires no prior knowledge of this minimum amplitude. It is designed for the case of practical relevance where the observations, and thus, the signals and noise, are two-dimensional random vectors, or, equivalently, complex random variables. Such observations can be the complex values returned by the standard I and Q decomposition of standard receivers in radar, sonar and telecommunication systems. In section 4.2, which addresses spectrum sensing, these complex values are those returned by the Discrete Fourier Transform (DFT) of the received signal. In the present section, we briefly present the MC-ESE since, in section 4.2, the fast-ESE will be compared to it. The interested reader will find further details in [9] and [6]. As in section 2.3, the signals are assumed to be less present than absent and to have finite energy.

To begin with, in the two-dimensional case, it follows from [14, Eq. 9.6.47, p. 377] that $\mathrm{I}_0(x) = {}_0\mathrm{F}_1(1; x^2/4)$ for every $x \in [0, \infty)$, where I_0 is the zeroth-order modified Bessel function of the first kind, so that $\xi(\rho) = \mathrm{I}_0^{-1}(e^{\rho^2/2})/\rho$ for any $\rho \in [0, \infty)$. The expression of Υ_0 simplifies as well when $d = 2$ and we have $\Upsilon_0(\tau) = 1 - \exp(-\tau^2/2)$ for $\tau \in [0, \infty)$. By taking into account that $\xi(0) = \sqrt{2}$ — since in the general d-dimensional case, $\xi(0) = \sqrt{d}$ [13] — we then estimate σ_0 by

$$\widetilde{\sigma_0} = \underset{\sigma \in [\sigma_{\min}, \sigma_{\max}]}{\operatorname{argmin}} \sup_{\ell \in \{1,\ldots,L\}} \left\{ \left| \frac{\sum_{k=1}^{m} \| Y_k \| \mathrm{I}(\| Y_k \| \leqslant \beta_\ell \sigma \sqrt{2})}{\sum_{k=1}^{m} \mathrm{I}(\| Y_k \| \leqslant \beta_\ell \sigma \sqrt{2})} - \sigma \frac{\Upsilon_1(\beta_\ell \sqrt{2})}{\Upsilon_0(\beta_\ell \sqrt{2})} \right| \right\}.$$

where $[\sigma_{\min}, \sigma_{\max}]$ is the same search interval as in Eq. (15). A minimization routine for scalar bounded non-linear functions, such as the `MATLAB` routine `fminbnd.m`, can be used for the computation of this estimate.

Although our choice to set $\mathsf{a} = 0$ is in contradiction with the asymptotic conditions involved in Eq. (7), it turns out that $\widetilde{\sigma_0}$ is a reasonably good estimate of

σ_0. However, the estimation can be improved by defining the MC-ESE

$$\widehat{\sigma_0} = \lambda \sqrt{\sum_{k=1}^{m} \|Y_k\|^2 \mathrm{I}(\|Y_k\| \leqslant \widetilde{\sigma_0}\sqrt{2}) / \sum_{k=1}^{m} \mathrm{I}(\|Y_k\| \leqslant \widetilde{\sigma_0}\sqrt{2})}. \tag{16}$$

where λ is some constant to choose with respect to the application. As proposed in [6], a reasonable choice for this constant is $\lambda = \sqrt{\Upsilon_0(\sqrt{2})/\Upsilon_2(\sqrt{2})} = 1.0937$.

As shown in [6, 7], the MC-ESE outperforms the standard Median Absolute Deviation (MAD) estimator for application to CES interception. In fact, the former is far more resilient than the latter when the number or the amplitudes of the outliers are too large. However, a mathematical justification of the MC-ESE remains an open issue. Moreover, the MC-ESE is computationally expensive because of the minimization routine it requires. In the next section, we then introduce the fast-ESE, whose computational load is far lesser than that of the MC-ESE and which can be justified in contrast to the MC-ESE, theoretically. It is also worth noticing that, in contrast with the MC-ESE again, the fast-ESE is not dedicated to the two-dimensional case.

3 The fast-ESE

In this section, we seek an estimator whose computational load is less than those of the estimators proposed above. This new estimator is obtained by simplifying the basic discrete cost of Eq. (12). We thus begin with some theoretical results allowing for choosing such a simpler discrete cost. Afterward, in section 3.2, we present the algorithm derived from these theoretical results. This algorithm is the "parameterized fast-ESE", so named because it depends on one single parameter, the so-called minimum signal to noise ratio (SNR). As explained below, depending on the application, this parameter can be chosen empirically. However, in section 3.3 and on the basis of our theoretical results, we present an heuristic approach and a few simulations on artificial and suitable signals to exhibit a seemingly reasonable value for this minimum SNR. The resulting version of the fast-ESE is called the universal fast-ESE because it is designed independently of any application. Two applications of the universal fast-ESE will illustrate its relevance in section 4.

3.1 Theoretical results

The first step toward a discrete cost simpler than that of Eq. (12) stems from the following remark about the ratio

$$\mathrm{M}_m(\sigma, T) = \sum_{k=1}^{m} \|Y_k\|^r \mathrm{I}(\|Y_k\| \leqslant \sigma T) / \sum_{k=1}^{m} \|Y_k\|^s \mathrm{I}(\|Y_k\| \leqslant \sigma T),$$

defined for every pair (σ, T) of positive real values. Basically, theorem 1 states that, when the minimum amplitude $\mathsf{a}(\Lambda)$ is large enough, $\mathrm{M}_m(\sigma_0, \theta(\mathsf{a}(\Lambda)/\sigma_0))$ is close to $\sigma_0^{r-s}\Psi(\mathsf{a}(\Lambda)/\sigma_0)$ where $\Psi(\rho) = \Upsilon_r(\theta(\rho))/\Upsilon_s(\theta(\rho))$ is defined for every $\rho \in [0,\infty)$. Defining the *minimum SNR* by $\varrho(\Lambda) = \mathsf{a}(\Lambda)/\sigma_0$, we can thus expect that $\mathrm{M}_m(\sigma_0, \theta(\varrho(\Lambda)))$ is actually close to $\sigma_0^{r-s}\Psi(\varrho(\Lambda))$. On the other hand, it follows from [15, p. 337, Eq. 3.326 (2)] that $\lim_{x\to\infty} \Upsilon_q(x) = 2^{\frac{q+d}{2}-1}\Gamma(\frac{q+d}{2})$ where Γ is the standard Gamma function. Therefore, from the asymptotic behavior of the thresholding function θ, the limit of $\Psi(\rho)$ exists when ρ tends to ∞ and is equal to

$$\Psi(\infty) = 2^{\frac{r-s}{2}} \Gamma\left(\frac{d+r}{2}\right) / \Gamma\left(\frac{d+s}{2}\right). \tag{17}$$

Remark 1 *Note that* $\Psi(\infty) = 0.7979$ *for* $d = 1$ *and* $\Psi(\infty) = 1.2533$ *for* $d = 2$.

According to the foregoing, $\mathrm{M}_m(\sigma_0, \theta(\varrho(\Lambda)))$ must be close to $\sigma_0^{r-s}\Psi(\infty)$ when $\varrho(\Lambda)$ is large enough. This is actually true in the sense specified by lemma 1 below. In this lemma, in the same way as we have defined the limit of a map of $\mathscr{M}(\Omega, \mathbb{R}^d)^{\mathbb{N}}$ into $\mathbb{R}$ when $\mathsf{a}(\Lambda)$ tends to ∞, we say that the limit of a given map f of $\mathscr{M}(\Omega, \mathbb{R}^d)^{\mathbb{N}}$ into $\mathbb{R}$ is $\ell \in \mathbb{R}$ when $\varrho(\Lambda)$ tends to ∞ and write that $\lim_{\varrho(\Lambda)\to\infty} f(\Lambda) = \ell$ if

$$\lim_{\rho\to\infty} \sup \left\{ |f(\Lambda) - \ell| : \Lambda \in \mathscr{M}(\Omega, \mathbb{R}^d)^{\mathbb{N}}, \varrho(\Lambda) \geqslant \rho \right\} = 0, \tag{18}$$

that is, if, for any positive real value η, there exists some $\rho_0 \in (0,\infty)$ such that, for every $\rho \geq \rho_0$ and every $\Lambda \in \mathscr{M}(\Omega, \mathbb{R}^d)^{\mathbb{N}}$ such that $\varrho(\Lambda) \geq \rho$, we have $|f(\Lambda) - \ell| \leqslant \eta$.

Lemma 1 *With the same notations and assumptions as theorem 1 and for any given thresholding function* θ, σ_0 *is the unique positive real number* σ *such that, for every* $\beta_0 \in (0,1]$,

$$\lim_{\varrho(\Lambda)\to\infty} \left\| \limsup_m \left| \mathrm{M}_m(\sigma, \beta\theta(\varrho(\Lambda))) - \Psi(\infty)\sigma^{r-s} \right| \right\|_\infty = 0 \tag{19}$$

uniformly in $\beta \in [\beta_0, 1]$.

PROOF: See appendix I. ■

On the basis of the previous result, the same type of reasoning as in [8, Sec. 3.1] leads to estimate the noise standard deviation σ_0 by seeking a possibly local minimum of

$$\underset{\sigma\in[\sigma_{\min},\sigma_{\max}]}{\operatorname{argmin}} \sup_{\ell\in\{1,\ldots,L\}} \left| \mathrm{M}_m(\sigma,\beta_\ell\theta(\varrho(\Lambda))) - \Psi(\infty)\,\sigma^{r-s} \right|, \tag{20}$$

where, as above, $\beta_\ell = \ell/L$ with $\ell = 1,2,\ldots,L$. We keep on working with the same search interval as above (see section 2.3) because the arguments of [8, Sec. 3.2] still apply here. The discrete cost of Eq. (20) is not really simpler than those of sections 2.3 and 2.4. However, because the discrete cost of Eq. (20) involves the constant $\Psi(\infty)$ instead of the ratio $\Psi(\rho)$, by setting $L = 1$ in Eq. (20), any solution in $\sigma \in [\sigma_{\min},\sigma_{\max}]$ to the equation

$$\mathrm{M}_m(\sigma,\theta(\varrho(\Lambda))) = \Psi(\infty)\,\sigma^{r-s} \tag{21}$$

can be considered as an estimate σ_0^* of σ_0. Lemma 2 below establishes the conditions for the existence of a solution in σ to equations such as Eq. (21). It also provides the value of this solution when these conditions are fulfilled. The fast-ESE simply derives from lemma 2.

Lemma 2 *With the notations above, define*

$$\mathrm{M}^*(k) = \begin{cases} \sum_{j=1}^{k} \|Y_{[j]}\|^r \Big/ \sum_{j=1}^{k} \|Y_{[j]}\|^s & \text{if } k \neq 0 \\ 0 & \text{if } k = 0, \end{cases} \tag{22}$$

for any $k = 0,1,2,\ldots$.

(i) *There exists a positive solution in σ to the equation*

$$\mathrm{M}_m(\sigma, T) = K\sigma^{r-s}, \tag{23}$$

where K and T are any two positive real numbers, if and only if there exists some integer $k \in \{1,2,\ldots,m\}$ such that

$$\|Y_{[k]}\| \leqslant \left(\mathrm{M}^*(k)/K\right)^{\frac{1}{r-s}} T < \|Y_{[k+1]}\|; \tag{24}$$

(ii) *If the inequalities in (24) are satisfied by some $k \in \{1,2,\ldots,m\}$, then $\left(\mathrm{M}^*(k)/K\right)^{\frac{1}{r-s}}$ is a solution in σ to Eq. (23).*

Remark 2 *In this statement, its proof and throughout the rest of the paper, we assume that the values $\| Y_k \|$ are all distinct, which is actually true with probability 1 because the probability distributions of the $\| Y_k \|$s are absolutely continuous with respect to Lebesgue's measure in $\mathbb{R}^d$. In any case, it would always be possible to consider the subset of all distinct values $\| Y_k \|$.*

PROOF: [Proof of lemma 2] We have $\mathrm{M}_m(\sigma, T) = \mathrm{M}^*(k(\sigma,\rho))$ where $k(\sigma,\rho)$ is the unique element of $\{0, 1, \ldots, m\}$ such that

$$\| Y_{[k(\sigma,\rho)]} \| \leqslant \sigma T < \| Y_{[k(\sigma,\rho)+1]} \| \tag{25}$$

with $Y_{[0]} = 0$ and $\| Y_{[m+1]} \| = \infty$.

Suppose the existence of σ such that Eq. (23) holds true. We then have $\mathrm{M}^*(k(\sigma,\rho)) = K\sigma^{r-s}$. Necessarily, $k(\sigma,\rho) \neq 0$ since $K\sigma^{r-s} > 0$. It follows that $\sigma = \left(\mathrm{M}^*(k(\sigma,\rho))/K\right)^{\frac{1}{r-s}}$. According to the inequalities of Eq. (25), we then obtain

$$\| Y_{[k(\sigma,\rho)]} \| \leqslant \left(\mathrm{M}^*(k(\sigma,\rho))/K\right)^{\frac{1}{r-s}} T < \| Y_{[k(\sigma,\rho)+1]} \|.$$

Thence, the existence of some element $k \in \{1, 2, \ldots, m\}$, actually equal to $k(\sigma,\rho)$, such that Eq. (24) holds true.

Conversely, suppose the existence of some integer $k \in \{1, 2, \ldots, m\}$ such that Eq. (24) holds true. Set $\sigma = (\mathrm{M}^*(k)/K)^{\frac{1}{r-s}}$. We then have $\| Y_{[k]} \| \leqslant \sigma T < \| Y_{[k+1]} \|$ so that $k(\sigma,\rho) = k$ and $\mathrm{M}_m(\sigma, T) = K\sigma^{r-s}$, which concludes the proof. ■

3.2 The parameterized fast-ESE

Let $Y = (Y_k)_{k\in\mathbb{N}}$ be some element of $\mathscr{M}(\Omega, \mathbb{R}^d)^{\mathbb{N}}$ such that $Y = \varepsilon\Lambda + X$ with $\Lambda = (\Lambda_k)_{k\in\mathbb{N}} \in \mathscr{M}(\Omega, \mathbb{R}^d)^{\mathbb{N}}$, $X = (X_k)_{k\in\mathbb{N}}$ is some d-dimensional AWGN with standard deviation σ_0 and $\varepsilon = (\varepsilon_k)_{k\in\mathbb{N}}$ is a sequence of random variables valued in $\{0, 1\}$. We make the same assumptions of practical interest as those made in section 2.3. We also assume that we are given a lower bound ρ for the minimum SNR $\varrho(\Lambda)$ and that ρ is large enough so that it is reasonable to estimate σ_0 by minimizing the discrete cost of Eq. (20). As suggested above, in order to alleviate the computational load, we restrict our attention to the case $L = 1$ so that our estimate is obtained by solving Eq. (21) in σ. We thus apply lemma 2 — with $K = \Psi(\infty)$ and $T = \theta(\rho)$ in Eq. (23) — to seek a solution $\sigma_0^* \in [\sigma_{\min}, \sigma_{\max}]$ for σ in Eq. (21). According to lemma 2 and the definition of the search interal $[\sigma_{\min}, \sigma_{\max}]$ (see section 2.3), this solution σ_0^* exists if and only if there exists $k \in \{k_{\min}, k_{\min}+1, \ldots, m\}$ such that $\| Y_{[k]} \| \leqslant \left(\mathrm{M}^*(k)/\Psi(\infty)\right)^{\frac{1}{r-s}} \theta(\rho) < \| Y_{[k+1]} \|$ where $k_{\min}$ is given by Eq. (14).

For the same reasons as those of section 2.3, the thresholding function θ will hereafter be ξ defined by Eq. (13) and we choose $r = 1$ and $s = 0$, for the same reasons as those given in section 2.3. Theoretically, values other than these ones could be chosen. However, some informal tests have not pinpointed more suitable values for r and s. Note also that our choice for r and s is computationally beneficial since the power $1/(r-s)$ is not used any longer to calculate the estimate (compare to lemma 2). The fast-ESE for minimum SNR ρ is thus specified as follows. We denote this algorithm by fast-ESE$_d(\rho)$ to emphasize that it is parameterized by ρ and applies to d-dimensional observations.

The parameterized fast-ESE$_d(\rho)$:

Input: A finite subsequence sequence $Y_1, Y_2, \ldots, Y_m$ of a sequence $Y = (Y_k)_{k \in n\mathbb{N}}$ of d-dimensional real random vectors satisfying assumptions **(A1-A4)** of theorem 1 with $\nu = 2$ and $\mathrm{p} = 0.5$.

Output: the estimate σ_0^* of the noise standard deviation under the assumption that ρ is a lower bound for the minimum SNR of the signals.

1. fix $k_{\min}$ according to Eq. (14),
2. if there exists a smallest integer $k \in \{k_{\min}, \ldots, m\}$ such that

$$\| Y_{[k]} \| \leqslant \mathrm{M}^*(k)\, \xi(\rho) / \Psi(\infty) < \| Y_{[k+1]} \| \tag{26}$$

   with

$$\mathrm{M}^*(k) = \begin{cases} \frac{1}{k} \sum_{j=1}^{k} \| Y_{[j]} \| & \text{if } k \neq 0 \\ 0 & \text{if } k = 0, \end{cases} \tag{27}$$

   for any $k = 0, 1, 2, \ldots, m$ and $\Psi(\infty) = \sqrt{2}\,\Gamma\left(\frac{d+1}{2}\right)/\Gamma\left(\frac{d}{2}\right)$, then, set

$$\sigma_0^* = \mathrm{M}^*(k)/\Psi(\infty) \tag{28}$$

3. otherwise, set $\sigma_0^* = +\infty$.

Remark 3 *Note that setting $L = 1$ in Eq. (12) amounts to seeking a solution in $\sigma \in [\sigma_{\min}, \sigma_{\max}]$ to the equation*

$$\frac{\sum_{k=1}^{m} \| Y_k \|^r \mathrm{I}(\| Y_k \| \leqslant \sigma\theta(\mathsf{a}/\sigma)))}{\sum_{k=1}^{m} Y_k \|^s \mathrm{I}(\| Y_k \| \leqslant \sigma\theta(\mathsf{a}/\sigma))} = \frac{\sigma \Upsilon_r(\theta(\mathsf{a}/\sigma))}{\Upsilon_s(\theta(\mathsf{a}/\sigma))}. \tag{29}$$

This equation is less simple than that of Eq. (21) since the coefficient in σ in the right hand side (rhs) of Eq. (29) depends on σ, which is not the case in Eq. (21). It is also worth noticing that setting $L = 1$ in Eq. (16) leads to seeking a solution in $\sigma \in [\sigma_{\min}, \sigma_{\max}]$ to the equation

$$\mathrm{M}_m(\sigma, \sqrt{2}) = \Psi(\sqrt{2})\,\sigma. \tag{30}$$

This equation is similar to Eq. (21) and lemma 2 could apply. However, we have not investigated further this possibility because the reasoning leading to the fast-ESE is far more elaborated. However, the present remark could be a clue for a general robust statistical signal processing framework embracing the MC-ESE and the fast-ESE and suggests some consistency between the MC-ESE and the fast-ESE when $L = 1$, beyond the fact that both derive from the same theoretical result.

As a consequence of the theoretical results presented in the paper, if the signals have a minimum signal to noise ratio above or equal to ρ, probabilities of presence less than or equal to $1/2$ and ρ and m are both large enough, the fast-ESE$_d(\rho)$ should return an accurate enough estimate σ_0^* of the noise standard deviation σ_0, with a computational load clearly less than that of the estimators of sections 2.3 and 2.4. However, in practice, how can we choose ρ, inasmuch as there is no particular reason for the signal norms to remain bounded away from zero? For a given application, it is thinkable to study the existence of a suitable value for ρ provided that we have a good enough model to generate synthetic noisy signals or, even better, a representative and sufficiently large database of real and suitably labeled observations. In the next section, we propose and test another solution for the determination of a value for ρ above which the fast-ESE$_d(\rho)$ performs well if the sample size m is large enough. This solution is experimental, depends on no particular application and leads to the notion of *universal* minimum SNR. The universal fast-ESE, described in the next section as well, is then simply the fast-ESE$_d(\rho)$ adjusted with ρ equal to this *universal* minimum SNR.

3.3 The universal minimum SNR and the universal fast-ESE

We are thus seeking a positive real value ϱ_u, called the universal minimum SNR, for which the so-called universal fast-ESE, that is, fast-ESE$_d(\varrho_u)$, is capable of achieving good estimates of the noise standard deviation in applications where the sample size is large enough, the signals have not necessarily large amplitudes and have possibly different probabilities of presence. If the minimum SNR can be calculated, the universal fast-ESE can be regarded as an alternative to the MAD estimator or the MC-ESE introduced above. Having yet no theoretical means to assess the behavior of fast-ESE$_d(\rho)$ through standard figures of merit such as the bias or the mean square error (MSE), we proceed experimentally to determine ϱ_u. Simulations aimed at determining ϱ_u should make it possible to compute the empirical bias and the empirical MSE when the signal probabilities of presence range over a discrete subset of $[0,1/2]$ and the minimum signal to noise ratio $\varrho(\Lambda)$ varies in some discrete subset of $(0,\infty)$. The difficulty is then to choose the signal distributions since the estimator is expected to perform whatever they are. The following heuristic reasoning provides an answer to this question.

Without loss of generality, suppose that $\sigma_0 = 1$. In this case, the minimum SNR is also the minimum amplitude or norm of the signals, which is convenient below for the sake of simplifying the presentation. Let ρ' be some positive real number such that $\rho' \geqslant \rho$. If the signals, represented by the sequence Λ, are such that $\varrho(\Lambda) \geqslant \rho' \geqslant \rho$, it can be expected that the larger $\varrho(\Lambda)$ with respect to ρ', that is, the larger the ratio $\varrho(\Lambda)/\rho'$, the better the performance of fast-ESE$_d(\rho)$ for the estimation of σ_0. This suggests that, for signals with SNRs larger than or equal to ρ', the least favorable case for fast-ESE$_d(\rho)$ is that of signals with norms equal to ρ'. For such signals, fast-ESE$_d(\rho)$ should thus return a rather accurate estimate of σ_0 if ρ'/ρ is large enough and the estimation should deteriorate when ρ' draws near ρ. When ρ' is lower than ρ and the signals have still norms equal to ρ', it can be expected that the closer ρ' to ρ, the lesser the performance of fast-ESE$_d(\rho)$. However, the estimation should improve as ρ'/ρ becomes small. According to the foregoing, for each positive ρ', our simulations concern signals Λ_k, $k = 1,2,\ldots$, that are elements of the sphere with radius ρ' and centered at the origin. Henceforth, we use the standard notation $\rho\mathrm{S}^{d-1}$ to designate the sphere with radius ρ and centered at the origin. Since there is no particular reason to favor any direction or angle for any Λ_k and the observations Y_k must be independent, the signals Λ_k are chosen iid, each being uniformly distributed on $\rho'\mathrm{S}^{d-1}$.

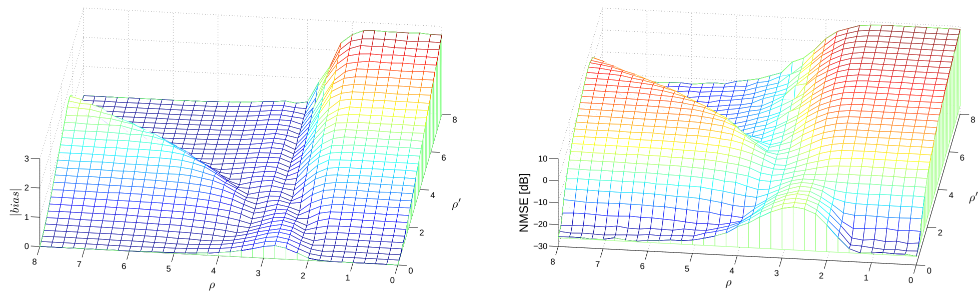
According to the foregoing, we proceeded as follows to determine a reasonable value ϱ_u for $d = 1,2$. Only the one- and the two-dimensional cases are considered because they are those encountered in many applications and especially in those of sections 4.1 and 4.2 below. Let SNRS = $\{0, 0.25, 0.5, 0.75, \ldots, 8\}$, PRIORS = $\{0.1, 0.2, \ldots, 0.5\}$ and SAMPLE_SIZES = $\{100, 200, \ldots, 500\}$. Given a pair (ρ,ρ') of elements of SNRS, some prior p in PRIORS and some sample size m in SAMPLE_SIZES, we generated N realizations of the sequence $Y_1 = \varepsilon_1\Lambda_1 + X_1, Y_2 = \varepsilon_2\Lambda_2 + X_2, \ldots, Y_m = \varepsilon_m\Lambda_m + X_m$ where, for $k = 1,2,\ldots,m$: the random vectors Λ_k are independent, each being uniformly distributed on $\rho'\mathrm{S}^{d-1}$; the random variables ε_k are independent, Bernouilli distributed each, valued in $\{0,1\}$ and such that $\mathrm{P}[\varepsilon_k = 1] = p$; the random vectors X_k are iid with $X_k \sim \mathcal{N}(0, \mathbf{I}_d)$.

The fast-ESE$_d(\rho)$ was then fed with each of these realizations so as to return N realizations of σ_0^*, the outcome of fast-ESE$_d(\rho)$. We then calculated the empirical bias $b_{m,\rho}(\rho', p)$ and the empirical MSE $Q_{m,\rho}(\rho', p)$ of σ_0^* for each tested values ρ' and m. On the basis of these empirical figures of merit for the accuracy of fast-ESE$_d(\rho)$, we determined, for each tested sample size m and each tested prior p, the values $\operatorname{argmin}_\rho \sum_{\rho'} b_{m,\rho}(\rho', p)$ and $\operatorname{argmin}_\rho \sum_{\rho'} Q_{m,\rho}(\rho', p)$ for ρ that respectively minimize the average empirical bias and average empirical MSE of fast-ESE$_d(\rho)$ over the tested amplitudes ρ' for the signals. These values are called the optimal minimum SNR with respect to the empirical bias and the optimal minimum SNR with respect to the empirical MSE, respectively. The results are summarized in tables 1 and 2. For instance, the empirical bias $b_{m,\rho}(\rho', p)$ and the empirical MSE $Q_{m,\rho}(\rho', p)$ in dB of σ_0^*, in the two-dimensional case, for $m = 100$ and $p = 0.5$, are those of figure 1; in this case, the optimum minimum SNRs with respect to the empirical bias and the empirical MSE were both found equal to 3.75. According to tables 1 and 2, the value of the optimum minimum SNRs with respect to the empirical bias and empirical MSE decrease with the probability of presence. The performance of the MAD estimator deteriorates in presence of too many outliers, that is, in our case dedicated to statistical signal processing, when the signal probabilities of presence exceed 0.3, roughly speaking. Thereby, **we chose the universal minimum SNR equal to** $\varrho_u = 4$, according to the experimental results of tables 1 and 2 for probabilities of presence larger than or equal to 0.3. Such a choice seems reasonable. However, depending on the application, other values for ϱ_u, especially for comparison with the MAD estimator when the signal have small probabilities of presence, could be considered.

Table 1: Optimal minimum SNR with respect to the empirical bias, given some sample size m and some signal probability of presence p.

	$m = 100$	$m = 200$	$m = 500$	$m = 1000$
$p = 0.1$	5.50	5.25	5.25	5.25
$p = 0.2$	4.75	4.75	4.75	4.75
$p = 0.3$	4.25	4.50	4.50	4.25
$p = 0.4$	4.00	4.00	4.00	4.00
$p = 0.5$	3.75	3.50	3.75	4.00

In the one-dimensional case, simulations of the same type as above led to the same type of results. In particular, it turned out that $\varrho_u = 4$ remains appropriate to adjust the fast-ESE in the one-dimensional case as well. Accordingly,



(a) Empirical bias (b) Empirical MSE

Figure 1: Empirical bias and MSE for the estimate σ_0^* returned by fast-ESE$_d(\varrho)$ when $d = 2$, the noise standard deviation is 1 and the signals are uniformly distributed on the sphere $\rho' \mathrm{S}^{d-1}$. The signal probabilities of presence all equal one half. The sample size is $m = 100$. For each pair (ρ, ρ'), the empirical bias and MSE are calculated over $N = 500$ independent tests involving m observations, each.

whether the dimension is 1 or 2, the fast-ESE$_d(\varrho_u)$ is henceforth called the universal fast-ESE.

At this stage, two remarks must be done, as an introduction to the next section. The first remark concerns the value found for ϱ_u. This value equals 4 and may seem quite large if ϱ_u is falsely interpreted as the sole SNR above which the universal fast-ESE can actually achieve a very good estimate of the noise standard deviation, whatever the signal distributions. In fact, the minimum SNR is actually the SNR above which the universal fast-ESE returns an accurate estimate of the noise standard deviation and below which this same estimator does not perform well, in the very specific case where the signal probabilities of

Table 2: Optimal minimum SNR with respect to the empirical MSE, given some sample size m and some signal probability of presence p.

	$m = 100$	$m = 200$	$m = 500$	$m = 1000$
$p = 0.1$	5.25	5.25	5.25	5.00
$p = 0.2$	4.75	4.75	4.50	4.50
$p = 0.3$	4.25	4.25	4.25	4.25
$p = 0.4$	4.00	4.00	4.00	4.00
$p = 0.5$	3.75	4.00	4.00	4.00

presence all equal one half and the signal norm distributions are Dirac masses. In the general case, much will actually depend on the signal norm distributions and the signal probabilities of presence.

3.4 Discussion on the robustness of the fast-ESE

The robustness of an estimator is characterized by its behavior in the "neighborhood" of a predefined model, i.e. the white Gaussian model in our case. To gain some insight, the theory of robust estimation defines several criteria of robustness. The influence function (IF) and the breakdown point are such very popular criteria.

The IF reflects the bias of any estimator $\hat{\Theta}$ at the underlying distribution F caused by the addition of a few outliers at point κ, standardized by the amount η of contamination, i.e.[3]

$$\mathrm{IF}_{\hat{\Theta}}(\kappa, F) = \lim_{\eta \downarrow 0} \frac{\hat{\Theta}\left((1-\eta)F + \eta\delta_\kappa\right) - \hat{\Theta}(F)}{\eta} \tag{31}$$

where δ_κ is the point-mass at κ. From the IF, several robustness measures can be defined such as

- the gross-error sensitivity: $\sup\limits_{\kappa}\left\{\left|\mathrm{IF}_{\hat{\Theta}}(\kappa, F)\right|\right\}$,
- the rejection point: $\inf\limits_{\zeta>0}\left\{\mathrm{IF}_{\hat{\Theta}}(\kappa, F) = 0, |\kappa| > \zeta\right\}$.

To get quantitative results on these measures, the closed-form expression of the IF is required. However, the fast-ESE belonging to the family of L-estimators, its IF depends on some quantile [5, p. 56] that is here indirectly determined by variable k of Eq. (26). Since k is dependent on the observation vector, it is therefore a random variable whose distribution has yet to be determined. Consequently, the IF expression cannot be derived. Although a quantitative analysis cannot be provided, qualitative conclusions can be drawn from what is known as the sensitivity curve. The sensitivity curve can be seen as a finite sample version of the IF [16, p. 43]. It measures the effect of different locations of an outlier κ on any estimate $\hat{\Theta}$ based on the sample $Y = Y_1, \cdots, Y_m$ and is defined as $SC_m(\kappa) = (m+1)\left(\hat{\Theta}(Y_1, \cdots, Y_m, \kappa) - \hat{\Theta}(Y_1, \cdots, Y_m)\right)$. Figure 2 shows the sensitivity curve of the universal fast-ESE and the IF of the MAD when the model distribution is Gaussian with $d = 1$. From this figure, it can first be conjectured that the gross-error sensitivity of the universal fast-ESE is finite and relatively small, which is expected from a robust estimator. Note that the gross-error sensitivity of the MAD is equal to 1.167, which is the smallest value that can be obtained for any scale estimator in the case of the normal distribution.

Moreover, the rejection point of the universal fast-ESE appears to be finite (≈ 2) in contrast to the MAD whose rejection point is infinite. Estimators with a finite rejection point are said to be redescending and are well protected against very large outliers. However, a finite rejection point can affect the efficiency of the estimator since samples near the tail of a distribution may be ignored. This finite rejection point also indicates that the universal fast-ESE can be used for detecting outliers or equivalently for detecting signals in white gaussian noise. Suppose the existence of a smallest integer $k \in \{k_{\min}, \ldots, m\}$ such that Eq. (26) holds true. Since the estimate of the noise standard deviation is then given by Eq. (28), it follows that, $\|Y_{[\ell]}\| \geqslant \sigma_0^* \xi(\rho), \ell \in \{k+1, \ldots, m\}$. By assumption, the signals have their norms larger than or equal to ρ. According to [13, Theorem VII-1], the thresholding test with threshold height $\sigma_0 \xi(\rho)$ is capable of detecting our signals with a probability of error upper bounded by a function of ρ, this function of ρ being also an upper bound for the probability of error of the *minimum probability of error* (MPE) test (see [17, Sec. II. B], among others, for the definition of the MPE test). This non trivial bound is sharp because it is attained when the signal has probability of presence equal to $1/2$ and has uniform distribution on $\rho \mathbb{S}^{d-1}$. Therefore, the fast-ESE$_d(\rho)$ can be regarded, and possibly used, as a detector of outliers, that is, under the assumptions considered above, as a detector of the signals that are either present or absent in noise. This detector thus decides that ε_ℓ equals 1 — and, thus, that Y_ℓ pertains to some signal — for every $\ell \geqslant k+1$ and decides that ε_ℓ equals 0 — and, thus, that Y_ℓ is noise alone — for every $\ell \leqslant k$.

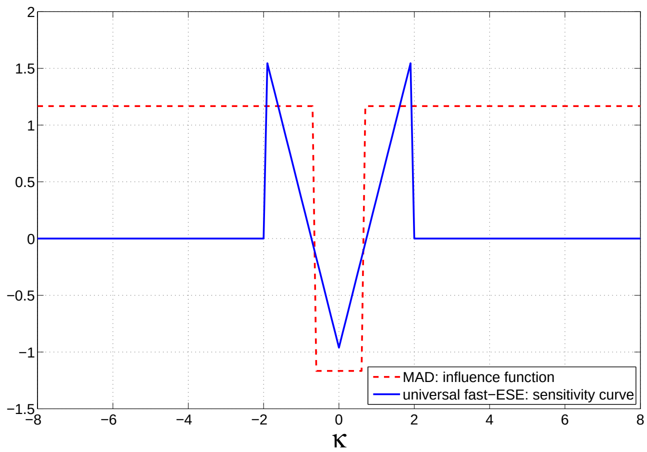


Figure 2: Sensitivity curve of the universal fast-ESE (m=10000) and influence function of the MAD with a one-dimensional Gaussian model.

Another measure of robustness is the breakdown point of an estimator. Several definitions for this measure are available in the literature [18]. In contrast to most scale estimators such as the MAD, the breakdown point of the universal fast-ESE depends on the actual realization of the observation vector Y so that it cannot easily be derived in the general case. This is once again due to the random nature of k in Eq. (26). Further theoretical studies have yet to be performed to address the derivation of the breakdown point.

4 Applications

4.1 WaveShrink

This application addresses the non-parametric estimation of a signal in the sense of [19]. The purpose is to recover an unknown deterministic function — the signal — from a noisy observation when noise is independent, additive, white and Gaussian. We are given a sequence of N observations $\boldsymbol{y} = \{y_i\}_{1 \leqslant i \leqslant N}$ such that $y_i = f(t_i) + e_i$ for $i = 1, 2, \ldots, N$, where f is the unknown function to recover, the random variables $\{e_i\}_{1 \leqslant i \leqslant N}$ are independent and identically distributed (iid), Gaussian with null mean and variance σ_0^2 so that $e_i \overset{\text{iid}}{\sim} \mathcal{N}(0, \sigma_0^2)$. The unknown deterministic function $f(\cdot)$ can be estimated as follows [19]. First, an orthonormal matrix $\mathscr{W}$ is applied to $\boldsymbol{y}$. The outcome of this transform is the sequence $c_i = d_i + x_i$, $i = 1, 2, \ldots, N$, with $\boldsymbol{c} = \{c_i\}_{1 \leqslant i \leqslant N} = \mathscr{W}\boldsymbol{y}$, $\boldsymbol{d} = \{d_i\}_{1 \leqslant i \leqslant N} = \mathscr{W}\boldsymbol{f}$, $\boldsymbol{f} = \{f(t_i)\}_{1 \leqslant i \leqslant N}$ and $\boldsymbol{x} = \{x_i\}_{1 \leqslant i \leqslant N} = \mathscr{W}\boldsymbol{e}$, $\boldsymbol{e} = \{e_i\}_{1 \leqslant i \leqslant N}$. The random variables $\{x_i\}_{1 \leqslant i \leqslant N}$ are iid and $x_i \sim \mathcal{N}(0, \sigma_0^2)$. The transform $\mathscr{W}$ is chosen so that most of the coefficients d_i, $i = 1, \ldots, N$, are small, even null, and only a few of them have large amplitude. Such a transform is said to achieve a sparse representation of the signal [20]. The wavelet transforms are generally regarded as sparse. Since the signal is characterized by only a few large coefficients, a non-linear filtering is applied to the coefficients returned by $\mathscr{W}$. On the one hand, this non-linear filtering forces to zero the small coefficients c_i because they are considered to derive from too small, or even null, components of the signal. On the other hand, it reduces the noise influence on the coefficients with amplitudes above the threshold because such coefficients are regarded as large enough to contain most of the information about the signal to estimate. This filtering stage is often performed by the so-called thresholding function $\delta_\lambda(\cdot)$ because of its well-known and desirable properties of smoothness and adaptation (see [21]). This thresholding function is defined by $\delta_\lambda(x) = x - \operatorname{sgn}(x)\lambda$ if $|x| \geqslant \lambda$ and $\delta_\lambda(x) = 0$, otherwise, where $\operatorname{sgn}(x) = 1$ (resp. -1) if $x \geqslant 0$ (resp. $x < 0$). The main role of the threshold λ is to distinguish large from small coefficients. Let $\widehat{\boldsymbol{d}} = \{\delta_\lambda(c_i)\}_{1 \leqslant i \leqslant N}$ stand for the outcome of the thresholding function, the estimate of $\boldsymbol{f}$ is then $\widehat{\boldsymbol{f}} = \mathscr{W}^{\top}\widehat{\boldsymbol{d}}$, where $\mathscr{W}^{\top}$ is the transpose, and thus, the inverse, of $\mathscr{W}$. Several thresholds have been proposed in the literature to improve the performance of WaveShrink by soft thresholding. The basic universal and minimax thresholds were introduced in [19]. The detection threshold was proposed in [10, 11]. Level-dependent thresholds, discussed in [11], are detection thresholds adjusted to each wavelet decomposition level, so as to adapt the soft thresholding to the decreasing of the amplitudes of the signal wavelet coefficients from one decomposition level to another. The performance of an image denoising method is generally evaluated by calculating the Peak Signal to Noise Ratio (PSNR). The PSNR is given by $\mathrm{PSNR}(\lambda) = 10\log_{10}\left(255^2 / r_{\lambda}(\boldsymbol{d}, \widehat{\boldsymbol{d}})\right)$ where λ is the threshold of the thresholding function and $r_{\lambda}(\boldsymbol{d}, \widehat{\boldsymbol{d}}) = \frac{1}{N}\mathbb{E}\|\boldsymbol{d} - \widehat{\boldsymbol{d}}\|^2 = \frac{1}{N}\sum_{i=1}^{N}\mathbb{E}\Big(d_i - \delta_{\lambda}(c_i)\Big)^2$ is the risk function or cost used to measure the accuracy of the estimate $\widehat{\boldsymbol{f}}$ of $\boldsymbol{f}$. According to [10, 11], if standard images corrupted by independent AWGN with known standard deviation σ_0 are denoised by WaveShrink with soft thresholding, the level-dependent thresholds achieve larger PSNRs and, thus, smaller risks, than the detection threshold, which, in turn, outperforms the minimax threshold, the universal threshold yielding generally the lesser PSNRs. The aforementioned thresholds all depend on the value of the noise standard deviation, which is often unknown in practice. An estimate can then be used instead of the actual value of the noise standard deviation. The MAD estimator is then standardly used to estimate σ_0 on the basis of the detail wavelet coefficients at the first decomposition level [19] since, due to the sparsity of the wavelet transform, only a few of these wavelet coefficients are actually large. The same reason can be evoked to justify the use of the universal fast-ESE. Therefore, the natural question that arises at this stage is to what extent the performance measurements of WaveShrink by soft thresholding are affected when the MAD estimate or the universal fast-ESE estimate are used instead of the exact value of σ_0 to adjust the shrinkage. To answer this question experimentally, we considered the standard 512×512 'Lena' image corrupted by independent AWGN with various values for the noise standard deviation. For each tested value of the noise standard deviation σ_0, we estimated σ_0 via the MAD and the universal fast-ESE applied to the detail coefficients of the first decomposition level. We then calculated the normalized empirical bias and normalized empirical MSE of these two estimates. The normalized empirical bias (resp. empirical MSE) are defined as the ratio between the empirical bias (resp. empirical MSE) and the actual value of σ_0. We also evaluated the PSNR after denoising by wavelet shrinkage based on the stationary wavelet transform (SWT) — for its appropriate properties in denoising [22] — and when the soft thresholding function is tuned with either the exact value of the noise standard deviation, its MAD estimate or the estimate returned by the universal fast-ESE.

Table 3 displays the absolute value of the normalized empirical bias and the value of the normalized empirical MSE of the MAD and the universal fast-ESE calculated over 100 trials. On the basis of these figures of merit, the universal fast-ESE outperforms the MAD estimator, even though the difference is not that tremendous. Table 4 presents the PSNRs obtained by wavelet shrinkage with soft thresholding, for the same values of the noise standard deviation as above and for the different ways to fix the threshold of the soft thresholding function. Once again, the universal fast-ESE performs slightly better than the MAD to yield PSNRs very close to those obtained when the noise standard deviation is known. For a given type of threshold, the results are very similar whether the exact value or the estimates of the noise standard deviation are used to adjust the thresholding function thanks to the good performance of the MAD and universal fast-ESE estimators.

Table 3: Absolute value of the normalized empirical bias and value of the normalized empirical MSE for the MAD and the universal fast-ESE when the standard Lena image is corrupted by independent AWGN with standard deviation σ_0.

	MAD		universal fast-ESE	
σ_0	|Bias|	MSE	|Bias|	NMSE
9	0.1533	0.0247	0.1047	0.0117
18	0.0539	0.0036	0.0296	0.0013
27	0.0268	0.0014	0.0002	0.0004

4.2 Spectrum sensing

In a cognitive radio context or for CES applications, spectrum sensing aims at monitoring the radio-frequency bands of interest to detect either communication systems or spectrum holes (i.e. idle frequency channels). In both cases, most of the detection algorithms rely on prior knowledge of the noise standard deviation. In these applications, detection usually results from the processing of a complex-valued observation represented in the time-frequency domain (i.e. at the output of a short-time Fourier transform), that is

$$Y_{k,n} = \varepsilon_{k,n}\Lambda_{k,n} + X_{k,n}, \tag{32}$$

where k is the time frame index and n the DFT bin number, $X_{k,n}$ is complex white Gaussian noise, $X_{k,n} \overset{\text{iid}}{\sim} \mathscr{CN}(0, 2\sigma_0^2)$ and $\Lambda_{k,n}$ is the received signal. This

Table 4: PSNRs after denoising by WaveShrink based on SWT when the soft thresholding function is tuned with either the exact value of the noise standard deviation, its MAD estimate or the estimate returned by the universal fast-ESE.

		Universal threshold	Minimax threshold	Detection threshold	Level-dependent thresholds
$\sigma_0 = 9$	known σ_0	29.30	30.61	32.08	33.62
	MAD estimate	28.71	29.99	31.46	33.15
	fast-ese estimate	28.89	30.18	31.65	33.30
$\sigma_0 = 18$	known σ_0	26.60	27.75	29.11	30.96
	MAD estimate	26.42	27.54	28.89	30.82
	fast-ese estimate	26.50	27.63	28.99	30.91
$\sigma_0 = 27$	known σ_0	25.27	26.26	27.49	29.42
	MAD estimate	25.21	26.19	27.41	29.37
	fast-ese estimate	25.27	26.26	27.49	29.41

signal is here modeled for the numerical applications by

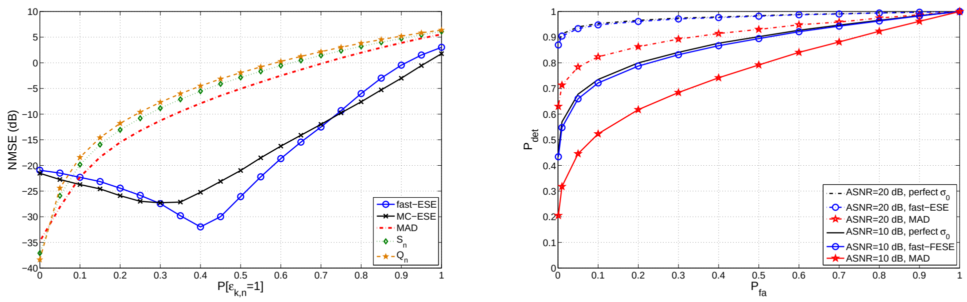
$$\Lambda_{k,n} = \sqrt{E_s}\, a_{k,n} H_{k,n}, \tag{33}$$

where E_s is the signal power, the $a_{k,n}$ represents the transmitted data symbol and $H_{k,n}$ is the propagation channel. For the sake of generality, the $a_{k,n}$'s are assumed to be iid, zero-mean and uniformly distributed with $\mathbb{E}[|a_{k,n}|^2] = 1$; $H_{k,n}$ is an iid Rayleigh fading channel in the frequency domain and a Gauss-Markov process in the time domain with $\mathbb{E}[H_{k,n} H^*_{k-1,n}] = 0.9$. The Average Signal to Noise Ratio (ASNR) is defined as $\text{ASNR} = 10\log_{10}\left(E_s/(2\sigma_0^2)\right)$.

Figure 3a compares the performance of the universal fast-ESE with that of the MC-ESE, the MAD and its alternatives S_n and Q_n depicted in [4][1]. These results are obtained in an "average scenario" where ASNR equals 10 dB and the observation is limited to 128 DFT bins and 16 time frames. Despite the theoretical ground of the universal fast-ESE that bounds the probability of the signal presence to 1/2, the NMSE of the estimated σ_0 is here plotted for $0 \le \mathrm{P}[\varepsilon_k = 1] \le 1$. This is motivated by the application context of spectrum sensing algorithms that can face situations where $\mathrm{P}[\varepsilon_k = 1] > 1/2$. As shown in [23], the occupancy rate of the radio-frequency spectrum varies form one band to another. CES systems mainly focus on military bands such as the 30-88 MHz,

[1]Note that the MAD and its alternatives are originally designed for real-valued random variables. However, since the real and imaginary parts of $Y_{k,n}$ are independent in our scenario, the noise standard deviation of $X_{k,n}$ is estimated on the observation vector made of the concatenation of the 2 dimensions of the $Y_{k,n}$'s.

225-400 MHz or 960-1240 MHz bands where the occupancy rate can be less than 10% in peace zones to more than 50% on theatres of operation. Cognitive radio systems mainly focus on TV and ISM bands where the occupancy rate is usually high (from 15% to 70% [23]). Figure 3a clearly shows that the universal fast-ESE and the MC-ESE largely outperform the MAD and its alternatives in the range of probability of presence relevant for spectrum sensing applications. This very good robustness is paid back by a poor efficiency (there is a 17 dB NMSE loss compared to Q_n for $\mathrm{P}[\varepsilon_k = 1] = 0$). This is consistent with the finite rejection point of the sensitivity curve observed in Figure 2. Figure 3a also indicates that the universal fast-ESE and the MC-ESE yield similar performance measurements. The universal fast-ESE is slightly better for $0.3 \leq \mathrm{P}[\varepsilon_k = 1] \leq 0.7$. This result advocates the use of the universal fast-ESE since this estimator is far less complex than the MC-ESE. It is also worth noticing the seemingly existence of an optimum for the NMSE of both estimators. This was not expected theoretically and deserves some attention in forthcoming work.



(a) (b)

Figure 3: Performance of the universal fast-ESE in a spectrum sensing context where $Y_{k,n}$ is a 16×128 matrix. (a) Normalized mean square error comparison of various robust noise standard deviation estimators with ASNR=10 dB. (b) Comparison of the receiver operating characteristics when the noise variance is perfectly known and when it is estimated via the universal fast-ESE and the MAD. The priors $\mathrm{P}[\varepsilon_{k,n} = 1]$ all equal $1/2$.

As detection is usually the most critical operation in spectrum sensing, the proposed estimator is indirectly evaluated in Figure 3b through the performance of a classical constant false alarm rate (CFAR) detector[2]. This figure compares the true detection rate (P_{det}) for various theoretical false alarm rates (P_{fa}) when the noise standard deviation is perfectly known and when it is estimated using

[2]CES or cognitive radio systems may use detector structures other than CFAR ones. This kind of detector structure is one of the most popular and is therefore used for illustration purposes.

the universal fast-ESE or the MAD. The decision on detection is made by comparing $|Y_{k,n}|^2$ to a positive threshold that aims at guaranteeing a specified false alarm rate. Given that noise is complex-valued and Gaussian, $|Y_{k,n}|^2/\sigma_0^2$ follows a chi-square distribution with 2 degrees of freedom when $\varepsilon_{k,n} = 0$. Therefore, when σ_0 is known, the detector decides that $\varepsilon_{k,n}$ equals 1 if $|Y_{k,n}|^2 > -2\sigma_0^2 \ln(P_{fa})$ and that $\varepsilon_{k,n}$ equals 0, otherwise. It is usual to summarize this decision-making on the value of $\varepsilon_{k,n}$ by writing $|Y_{k,n}|^2 \underset{\varepsilon_{k,n}=0}{\overset{\varepsilon_{k,n}=1}{\gtrless}} -2\sigma_0^2 \ln(P_{fa})$. When σ_0 is estimated by the universal fast-ESE or the MAD, we replace σ_0 by its estimate. Figure 3b confirms the benefit of the universal fast-ESE for spectrum sensing applications. In the example given, the MAD over-estimates σ_0 so that the true detection rate is far below the one obtained with a perfect knowledge of σ_0. This has to be compared with the detection rate of the universal fast-ESE, which is very similar to the ideal one.

5 Conclusion and perspectives

On the basis of [8, Theorem 1], a new robust estimate of the noise standard deviation has been proposed for problems of practical interest where d-dimensional observations involving signals with unknown distributions must be processed so as to extract knowledge about these same signals. This estimate is the fast-ESE. In contrast with the standard MAD estimator, which is limited to the one-dimensional case, the fast-ESE applies whatever the value of d. Once the sole parameter, on which the computation of this estimate depends, is fixed according to a few simulation results, the resulting universal fast-ESE has been shown to perform very well in two rather natural applications, namely, image denoising by wavelet shrinkage and spectrum sensing. The theoretical derivation of the fast-ESE and the universal fast-ESE is made under weak sparsity hypotheses that bound our lack of prior knowledge about the signal distributions and probabilities of presence.

The fast-ESE and the universal fast-ESE raise several questions and open new prospects in robust statistical signal processing. To begin with, further work is required to better understand the behaviour of the universal fast-ESE. In this respect, it could be relevant to take into account [8, Theorem 1, statement (ii)]. Indeed, the present paper as well as [6, 7] are based on [8, Theorem 1, statement (i)], which concerns the case of big signal amplitudes, whereas [8, Theorem 1, statement (ii)] states that the noise standard deviation is also the solution of a limit equation when signals have very small amplitude and priors upper-bounded by some value in $[0,1)$. Such a result could perhaps be exploited to justify the good performance of the MC-ESE and the universal fast-

ESE, even when the signal amplitudes or the signal SNRs are not big. It would also be helpful to characterize the probability distributions of the norms of the signals in the presence of which the universal fast-ESE is capable of estimating accurately the noise standard deviation. A theoretical analysis of the bias, MSE, efficiency and consistency of the fast-ESE and the universal fast-ESE is desirable so as to justify or improve the experimental considerations of the paper.

We also plan to study to what extent the MC-ESE, fast-ESE and universal fast-ESE can be cast into a unified framework, in connection with standard theory of robust statistics. A study of the several figures-of-merit that are standardly used to characterize the robustness of an estimator could then be undertaken. In this respect and according to the contents of this paper, we can make the following remarks. For one-dimensional observations, the fast-ESE and the MAD are very similar. Indeed, for $d = 1$, we have $\Psi(\infty)^{-1} = 1.2533$ (see remark 1), which is close to the value given to b in Eq. (1) for the Gaussian case. Besides, $\mathrm{M}^*(k)$ in Eq. (28) is analoguous to the median value involved in Eq. (1). Finally, remark 3 has already pinpointed a similarity between the MC-ESE and the fast-ESE. It is thus thinkable that the MAD, the fast-ESE, and even the MC-ESE, derive from a more general robust estimator, possibly based on order statistics. It could also be asked whether choosing $L \geqslant 2$ instead of $L = 1$ in Eq. (19) of lemma 1 could yield alternative estimators and to what extent these alternative estimators could also relate to the MAD and the MC-ESE.

Basically, the fast-ESE is an outlier detector. This has already been enhanced in section 3.4. However, the capability of the fast-ESE to detect outliers has not yet been exploited, even in section 4.2 where the detection is performed via a very standard approach. The link between the problem of detecting outliers or signals and the estimation problem addressed by the MC-ESE, the fast-ESE and the universal fast-ESE should deserve some attention.

Appendix I
Proof of lemma 1

The fact that σ_0 satisfies Eq. (19) is a straightforward consequence of Eqs. (3) and (7). The fact that σ_0 is actually the unique positive real satisfying Eq. (19) is proved by mimicking the proof in [8, Appendix A.5] with some slight simplification.

Assume the existence of two positive real numbers $\sigma_1 \geq \sigma_2 > 0$ that both satisfy Eq. (19). Let β_1 abd β_2 two elements of $(0,1)$ such that $\beta_1\sigma_2 = \beta_2\sigma_1$. We have

$$
\begin{aligned}
&\left|\sigma_1^{r-s} - \sigma_2^{r-s}\right| \Psi(\infty) \\
&\quad \leqslant \left| \frac{\sum_{k=1}^{m} \|Y_k\|^r \mathrm{I}_{[0,\beta_2\sigma_1\theta(\varrho(\Lambda))]}(\|Y_k(\omega)\|)}{\sum_{k=1}^{m} \|Y_k\|^s \mathrm{I}_{[0,\beta_2\sigma_1\theta(\varrho(\Lambda))]}(\|Y_k(\omega)\|)} - \sigma_1^{r-s}\Psi(\infty) \right| \\
&\qquad + \left| \frac{\sum_{k=1}^{m} \|Y_k\|^r \mathrm{I}_{[0,\beta_1\sigma_2\theta(\varrho(\Lambda))]}(\|Y_k(\omega)\|)}{\sum_{k=1}^{m} \|Y_k\|^s \mathrm{I}_{[0,\beta_1\sigma_2\theta(\varrho(\Lambda))]}(\|Y_k(\omega)\|)} - \sigma_2^{r-s}\Psi(\infty) \right|
\end{aligned}
$$

for any given pair $(m,\omega) \in \mathbb{N} \times \Omega$. From this inequality, we derive that

$$
\begin{aligned}
&|\sigma_1^{r-s} - \sigma_2^{r-s}|\Psi(\infty) \\
&\quad \leqslant \left\| \limsup_{m\to\infty} \left| \mathrm{M}_m(\sigma_1, \beta_2\theta(\varrho(\Lambda))) - \Psi(\infty)\,\sigma_1^{r-s} \right| \right\|_\infty \\
&\qquad + \left\| \limsup_{m\to\infty} \left| \mathrm{M}_m(\sigma_2, \beta_1\theta(\varrho(\Lambda))) - \Psi(\infty)\,\sigma_2^{r-s} \right| \right\|_\infty .
\end{aligned}
$$

Since σ_1 and σ_2 are both assumed to satisfy Eq. (19), then, by choosing $\varrho(\Lambda)$ large enough, it follows from Eq. (3) that the rhs in the inequality above can be rendered arbitrarily small. Therefore, we conclude that $\sigma_1 = \sigma_2$.

References

[1] R. Poisel, *Introduction to Communication Electronic Warfare Systems*, Artech House Publishers, 2002.

[2] I. F. Akyildiz and S. Mohanty W-Y. Lee, M. C. Vuran, "NeXt generation/dynamic spectrum access/cognitive radio wireless networks: A survey," *Computer Networks*, pp. 2127–2159, 2006.

[3] F.R. Hampel, "The influence curve and its role in robust estimation," *Journal of the American Statistical Association*, vol. 69, no. 346, pp. 383–393, Jun. 1974.

[4] P. J. Rousseeuw and C. Croux, "Alternatives to the Median Absolute Deviation," *Journal of the American Statistical Association*, vol. 88, no. 424, pp. 1273–1283, Dec. 1993.

[5] P.J. Huber and E.M. Ronchetti, *Robust Statistics, second edition,* John Wiley and Sons, 2009.

[6] F.-X. Socheleau, D. Pastor, and A. Aissa-El-Bey, "Robust statistics based noise variance estimation: Application to wideband interception of non-cooperative communications," *IEEE Trans. Aerosp. Electron. Syst.*, 2011.

[7] F.-X. Socheleau, D. Pastor, and A. Aissa-El-Bey, "A sparse model for robust noise variance estimation and application to wideband CES interception," Tech. Rep., Lab-STICC UMR CNRS 3192 - technologies de l'information, de la communication et de la connaissance (Institut Télécom-Télécom Bretagne), SC - Dépt. Signal et Communications (Institut Télécom-Télécom Bretagne), 2009.

[8] D. Pastor, "A theoretical result for processing signals that have unknown distributions and priors in white Gaussian noise," *Computational statistics and data analysis,* vol. 52, pp. 3167–3186, 2008.

[9] D. Pastor and A. Amehraye, "Algorithms and Applications for Estimating the Standard Deviation of AWGN when Observations are not Signal-Free," *Journal of Computers,* vol. 2, no. 7, Sept. 2007.

[10] G. Mercier A.M. Atto, D. Pastor, "Detection threshold for non-parametric estimation," *Signal, Image and Video Processing, Springer,* vol. 2, pp. 207–223, 2008.

[11] D. Pastor and A.M. Atto, *Wavelet shrinkage: from sparsity and robust testing to smooth adaptation,* To appear in Fractals and Related Fields, Eds. : J. Barral & S. Seuret, Birkhaüser, 2010.

[12] N. N. Lebedev, *Special Functions and their Applications,* Prentice-Hall, Englewood Cliffs, 1965.

[13] D. Pastor, R. Gay, and A. Gronenboom, "A sharp upper bound for the probability of error of likelihood ratio test for detecting signals in white gaussian noise," *IEEE Trans. Inf. Theory,* vol. 48, no. 1, pp. 228–238, Jan. 2002.

[14] M. Abramowitz and I. Stegun, *Handbook of Mathematical Functions. Ninth printing.,* Dover Publications Inc., New York., 1972.

[15] I.S. Gradshteyn and I.M. Ryzhik, *Table of integrals, series and products, 7th edition,* Edited by Allan Jeffrey and Daniel Zwillinger, Academic Press, Elsevier, 2007.

[16] F. R. Hampel, E.M. Ronchetti, P.J. Rousseeuw, and W.A. Stahel, *Robust Statistics: the Approach based on Influence Functions*, John Wiley and Sons, New York, 1986.

[17] H. V. Poor, *An Introduction to Signal Detection and Estimation*, Springer-Verlag, New York, 1994.

[18] D.L. Donoho and P.J. Huber, "The notion of breakdown point," *In A Festschrift for Erich L. Lehmann*, pp. 157–184, 1983.

[19] D. L. Donoho and I. M. Johnstone, "Ideal spatial adaptation by wavelet shrinkage," *Biometrika*, vol. 81, no. 3, pp. 425–455, Aug. 1994.

[20] I. M. Johnstone, "Wavelets and the theory of non-parametric function estimation," *Journal of the Royal Statistical Society*, vol. A, no. 357, pp. 2475–2493, 1999.

[21] D.L. Donoho, "De-noising by soft-thresholding," *IEEE Trans. Inf. Theory*, vol. 41, no. 3, pp. 613–627, May 1995.

[22] S. Mallat, *A wavelet tour of signal processing, second edition*, Academic Press, 1999.

[23] "Shared Spectrum Company," `http://www.sharedspectrum.com/measurements/`.

www.telecom-bretagne.eu

Campus de Brest
Technopôle Brest-Iroise
CS 83818
29238 Brest Cedex 3
France
Tél. : + 33 (0)2 29 00 11 11
Fax : + 33 (0)2 29 00 10 00

Campus de Rennes
2, rue de la Châtaigneraie
CS 17607
35576 Cesson Sévigné Cedex
France
Tél. : + 33 (0)2 99 12 70 00
Fax : + 33 (0)2 99 12 70 19

Campus de Toulouse
10, avenue Edouard Belin
BP 44004
31028 Toulouse Cedex 04
France
Tél. : +33 (0)5 61 33 83 65
Fax : +33 (0)5 61 33 83 75

Imprimé à Télécom Bretagne
Dépôt légal : juin 2010
ISSN : 1255-2275